

A Compact Linear Programming Relaxation for Binary Sub-modular MRF

Junyan Wang and Sai-Kit Yeung

Singapore University of Technology and Design, Singapore

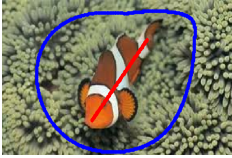
Abstract. We propose a novel compact linear programming (LP) relaxation for binary sub-modular MRF in the context of object segmentation. Our model is obtained by linearizing an l_1^+ -norm derived from the quadratic programming (QP) form of the MRF energy. The resultant LP model contains significantly fewer variables and constraints compared to the conventional LP relaxation of the MRF energy. In addition, unlike QP which can produce ambiguous labels, our model can be viewed as a quasi-total-variation minimization problem, and it can therefore preserve the discontinuities in the labels. We further establish a relaxation bound between our LP model and the conventional LP model. In the experiments, we demonstrate our method for the task of interactive object segmentation. Our LP model outperforms QP when converting the continuous labels to binary labels using different threshold values on the entire Oxford interactive segmentation dataset. The computational complexity of our LP is of the same order as that of the QP, and it is significantly lower than the conventional LP relaxation.

Keywords: Binary submodular MRF, Graph cuts, Linear programming relaxation, total variation, object segmentation

1 Introduction

Markov Random Field (MRF) is a fundamental model for various computer vision tasks. In an MRF model, an MRF energy is to be minimized in order to find an optimal solution to the task. Minimizing general MRF energies is NP-hard [13]. Nevertheless, it has been shown that a particular MRF energy can be minimized efficiently and exactly by using max-flow/min-cut algorithms, i.e. the graph cuts [9,3] and such MRF energy is known as the binary sub-modular MRF energy. A typical problem modeled by binary sub-modular MRF is object segmentation [2]. Accordingly, we also present our work in the context of object segmentation in this paper.

Recently, graph cuts has been criticized for its rigidity in modeling [16,1] the inconveniences caused in its parallelization [1]. The linear programming (LP) relaxation model has been promoted, since it offers a more flexible and parallelizable substitute to graph cuts for minimizing binary sub-modular MRF energies [16,1]. We will use LP relaxation model or LP model interchangeably if there is no risk of confusion. The conventional LP relaxation model [16,1] was obtained by relaxing the l_1 -norm pairwise potential in the binary sub-modular MRF. Bhusnurmath and Taylor [1] proved that the solution by the LP model is identical to the solution by graph cuts. They also argued that the most apparent benefit of LP model over graph cuts is that it allows for



Input image with seeds

Conventional LP [16]

QP [8,24]

Our method

Fig. 1: Conventional Linear programming (LP) [16] is computationally heavy but it preserves discontinuity of the labels at the boundary. Quadratic programming (QP) [8,24] has much lower computational complexity but often produces over-smooth labels at the boundary. We propose a novel compact LP relaxation that is capable of preserving discontinuities in the labels but with computational complexity in the same order of QP.

convenient parallelization, whereas the existing parallel implementations of graph cuts are based on heuristics, and their effectiveness can only be validated experimentally [18,23]. Therefore, the LP model can benefit more from the recent and future development on computing devices [7], and it can be more suitable for future applications.

The computational advantage of the LP relaxation model for the binary sub-modular MRF has not been fully revealed in the past. The existing LP relaxation model [16,1] contains a large amount of auxiliary variables and constraints, which will correspond to very large computational complexity [28]. It has been reported in [16] that it takes 10 minutes to compute a solution to the LP problem for an image of 300×300 pixels on a machine with Intel P4(2.8Ghz) CPU and 1G RAM memory, and it has also been shown in [1] that the computation time for solving the LP problem on a standard GPU is the same as that required by graph cuts for the same task on a standard CPU. This motivates us to explore the possibilities of reducing the number of auxiliary variables and constraints, i.e. the size of the LP problem, in order to improve the efficiency.

In contrast to the LP model, the computational complexity of the quadratic programming (QP) relaxation for the binary sub-modular MRF energy proposed in [8,24] is much smaller than that required by the conventional LP model. However, the QP model may produce over-smooth ambiguous labels at the boundaries and this may cause incorrect segmentation. Example continuous labels produced by conventional LP and QP are shown in Fig. 1, in which the solution by conventional LP is clean and more desirable than that by QP. This motivates us to leverage the benefits of both the QP and LP relaxations to achieve efficient and discontinuity-preserving labeling.

In this paper, we propose a novel compact LP relaxation for binary sub-modular MRF. Our LP relaxation is obtained by linearizing a novel l_1^+ -norm minimization problem that is derived from the QP relaxation model [8,24]. The complexity of the algorithm for solving the proposed LP problem is of the same order of the corresponding QP model, and it is significantly smaller than that of the conventional LP. According to our theoretical analysis, the derived new l_1^+ -norm minimization is strongly related to the conventional LP which is actually a total-variation minimization problem. Thus, it should be able to preserve discontinuities in labels, while the QP model can over-diffuse the labels. In the experiments, our method also produces segmentation results comparable to those of conventional LP while being more desirable than those of the

QP. Our method outperforms QP for all most any threshold values used for binarizing continuous labels on the entire Oxford object segmentation dataset. To summarize, our LP relaxation can be more desirable than the conventional LP and QP relaxations, if the computational cost of the conventional LP is unacceptable and the quality of the solution to QP is not sufficiently satisfactory. At last, we would like to mention that we focus on the binary sub-modular MRF in this work. Hence, our method should not be confused with the recent developments on the approximate solutions to general MRF models [13,14,11].

2 Background

2.1 The MRF model for object segmentation

The general model of object segmentation considered in this paper is the Potts model that contains two terms: an object-background region model term (data term) $\mathbf{A}(\cdot)$, a.k.a. the unary potential, and a boundary model term $\mathbf{B}(\cdot)$, a.k.a. the pairwise potential. The most general form of this model can be written as follows:

$$\min_{\mathcal{L}} \mathbf{A}(\mathcal{L}, \theta) + \lambda \mathbf{B}(\mathcal{L}), \quad (1)$$

where \mathcal{L} is a label vector corresponding to all pixels/superpixels, an element of \mathcal{L} is 1 if the pixel/superpixel is from the object, or 0 otherwise. In this paper, we consider superpixel henceforth. θ is the set of parameters for the region model, and λ is the penalty coefficient.

The object-background region model term is often defined by the probability models of the image values in the respective regions, as shown in the Boykov-Jolly model [2]. Given such model, we are able to find the optimal segmentation that corresponds to the highest probability.

The region model alone is generally insufficient for locating the object boundary accurately. Hence, a boundary model term, i.e. the pairwise potential, is often used. The general form of the boundary model term $\mathbf{B}(\cdot)$ can be written as

$$\mathbf{B}(\mathcal{L}) = \sum_{(i,j) \in \mathcal{E}} B_{ij} |\mathcal{L}_i - \mathcal{L}_j|^p, \quad (2)$$

where i and j are the indices of pixels/superpixels, $B_{ij} = \frac{1}{1 + \{\|I_i - I_j\|^2\}} + c$, I_i, I_j are the image values at the i - and j -th pixel/superpixel in the image. \mathcal{E} is a neighborhood system and p is either 1 or 2 in this paper. We will elaborate on the choice of the value of p in this paper. Note that the weight B_{ij} contains two parts. One is $\frac{1}{1 + \{\|I_i - I_j\|^2\}}$ which encourages a discontinuity at image edges, and the other is a constant c that imposes smoothness to the resultant boundary. The rationale of this form of weights in pairwise potential can be found in the literature of Markov Random Field based segmentation models [17]. The constant weight in the latter part is related to the curve-shortening flow and is more frequently used in the active contour models [12,6].

2.2 LP relaxation of l_1 -norm minimization for segmentation

The region model \mathbf{A} is often formulated linear in the label vector, and it is relatively easy to handle. The complexity of the optimization for the MRF model is often determined by the boundary model term. It has been pointed out that the minimization of the boundary model term is equivalent to an l_1 -norm minimization problem in [16] and [1] separately and individually. It is further shown in [16] and [1] that the l_1 -norm minimization problem can be solved by LP, and it is proven in [1] that the solution to the LP problem in [1] converges to either 0 or 1 without any external prodding. Careful readers will find that the linear constraints in the LP relaxations in [16] and [1] are slightly different. Without resorting to a rigorous proof, it is easy to see the equivalence between them by setting the RHS of the linear constraints in [1] to be close to zeros.

In this paper, we shall adopt the LP in [16] because its derivation is straightforward. The LP model can be rewritten as follows:

$$\begin{aligned} \min_{\mathcal{L}} \quad & \sum_j A_i \mathcal{L}_i + \sum_{(i,j) \in \mathcal{E}} B_{ij} X_{ij} \\ \text{s.t.} \quad & \forall (ij) \in \mathcal{E}, -X_{ij} \leq \mathcal{L}_i - \mathcal{L}_j \leq X_{ij} \\ & \forall i, 0 \leq \mathcal{L}_i \leq 1, \forall (i, j) \in \mathcal{E}, 0 \leq X_{ij}. \end{aligned} \quad (3)$$

where $\mathbf{A} = \sum_j A_i \mathcal{L}_i$, and the variable X_{ij} is an auxiliary variable induced by the linearization process. A drawback of this well-known LP formulation is that it requires a large number of auxiliary variables and constraints. Suppose there are N elements to be labeled, then there can be *as many as* $N + N \times N$ variables and $N + 2N \times N$ linear constraints, which is the worst case. Here, we briefly review the result of complexity analysis for LP. The number of variables, often denoted as n , and the number of constraints, denoted as m , are the main characteristics of the computational complexity of LP. The computational complexity of LP is known as $O(n^3)$ [28] and when n is fixed the complexity is $O(m)$ [19]. As a result, the computational complexity for solving the above LP problem is $O(N^6)$, and the computational cost can be high, which has been witnessed in [16].

Another interesting LP framework of segmentation has been reported in [22], in which Schoenemann *et al.* proposed an alternative linear formulation of smoothness and curvature priors in segmentation and inpainting.

In this paper, we are interested in the pairwise potential in MRF which is used in almost all the papers on segmentation based on MRF.

3 MRF and norms minimization

In this section, we revisit the connections between MRF and l_1 - and l_2 -norm minimization.

3.1 From MRF to l_1 -norm minimization.

The boundary potential defined in Eq. (2) can be reformulated as an l_1 -norm minimization problem when $p = 1$ in Eq. (2), which is conventional in graph cuts [3]. This is

written as follows:

$$\mathbf{B}_{l_1}(\mathcal{L}) = \sum_{(i,j) \in \mathcal{E}} E_{ij} B_{ij} |\mathcal{L}_i - \mathcal{L}_j| = \|\text{diag}(\mathbf{w})\mathbf{D}\mathcal{L}\|_{l_1}, \quad (4)$$

where $E_{ij} = 1$ if i and j are neighbors and $E_{ij} = 0$ otherwise. B_{ij} is defined previously. $[\text{diag}(\mathbf{w})]^{N^2 \times N^2}$ is the diagonal matrix composed of \mathbf{w} and $w_k = B_{ij}$, if $k = i + (j - 1) \times N$. \mathbf{D} is an incidence matrix defined as follow:

$$[D]_{ij}^{N^2 \times N} = \begin{cases} 1, & \text{if } j = (i \bmod N) \\ -1, & \text{if } (i \bmod N, j) \in \mathcal{E} \end{cases} \quad (5)$$

We shall call $\mathbf{B}_{l_1}(\mathcal{L})$ the l_1 -norm boundary term henceforth. This reformulation has been reported previously [24,16], and it was immediately considered as an LP problem in [16].

3.2 From MRF to l_2 -norm minimization.

In addition, the boundary potential can also be reformulated as an l_2 -norm minimization problem when $p = 2$. This is shown as follows:

$$\begin{aligned} \|\text{diag}(\mathbf{w})\mathbf{D}\mathcal{L}\|_{l_2} &= \langle \text{diag}(\mathbf{w})\mathbf{D}\mathcal{L}, \text{diag}(\mathbf{w})\mathbf{D}\mathcal{L} \rangle^{\frac{1}{2}} \\ &= \left(\sum_k w_k^2 (\mathbf{d}_k \mathcal{L})^2 \right)^{\frac{1}{2}} \\ &= \left(\sum_{(i,j) \in \mathcal{E}} B_{ij}^2 (\mathcal{L}_i - \mathcal{L}_j)^2 \right)^{\frac{1}{2}} \\ &\Leftrightarrow \sum_{(i,j) \in \mathcal{E}} B_{ij}^2 (\mathcal{L}_i - \mathcal{L}_j)^2 \end{aligned} \quad (6)$$

where \mathbf{d}_k is the k^{th} row vector of \mathbf{D} , and we may define:

$$\mathbf{B}_{l_2}(\mathcal{L}) = \|\text{diag}(\mathbf{w})\mathbf{D}\mathcal{L}\|_{l_2}, \quad (7)$$

and we shall call $\mathbf{B}_{l_2}(\mathcal{L})$ the l_2 -norm boundary term henceforth.

In the last line of Eq. (6), we omit the square-root operation since it is a scaling of the value without changing the optimality of the solutions and hence we use “=” instead of “ \Leftrightarrow ”. This quadratic form is exactly the l_2 -norm minimization model in [24], and it is known as the random walker model. The rationale of this model has been well justified for the model with continuous label values.

The relationship between l_1 -norm boundary term and l_2 -norm boundary term. A canonical relationship between the two terms can be characterized by the following classic inequalities [26]:

$$\frac{1}{\sqrt{n}} \|\mathbf{x}\|_{l_1} \leq \|\mathbf{x}\|_{l_2} \leq \|\mathbf{x}\|_{l_1}, \quad (8)$$

The above holds for any vector $\mathbf{x} \in \mathbb{R}^n$. In our case, $\mathbf{x} = \text{diag}(\mathbf{w})\mathbf{D}\mathcal{L}$ and $n = N \times N$. This relation between l_1 -norm and l_2 -norm is known as the *equivalence of norms*.

3.3 Comparing l_1 -norm minimization with l_2 -norm minimization

The l_2 -norm minimization problem is quadratic in the label variable, and it often results in smooth labels at the object boundary, which may cause ambiguity in the boundary location. In contrast, the original l_1 -norm minimization will offer clearly distinct labels. See Fig. 1 for one example. A similar problem was identified about 20 years ago in image denoising. It was observed that the minimization of square of image gradients will result in blurry edges. This leads to the invention of the celebrated ROF total-variation minimization model for denoising [21]. It has already been pointed out that the l_1 -norm minimization in our context corresponds to total variation minimization [5]. Likewise, the l_2 -norm minimization corresponds to the problem of minimization of square of gradients in the context of denoising.

A discontinuity will cause finite total variation, while it yields infinitely large square of gradient. Hence, the discontinuities can be preserved during the minimization of total variation, while they are allowed by minimization of square of gradient. Hence, the total-variation minimization is more preferable to the minimization of square of gradient, as discontinuities are prevalent in images.

In segmentation, the solution from l_2 -norm minimization may also become over-smooth and therefore ambiguous at the boundaries. This can affect the accuracy of boundary locating in the segmentation. Accordingly, we also expect the solution of our model to contain sharp discontinuities, which is often allowed by the l_1 -norm minimization [21,25,4].

4 A compact LP model for object segmentation

4.1 A compact LP relaxation of l_1 -norm minimization by factorizing l_2 -norm

In the following, we will show that a new l_1 -norm, which is induced by factorizing the l_2 -norm boundary term in Eq.(6), can lead to a more compact LP problem with significantly less computational complexity compared to the original LP problem.

First, we rewrite the l_2 -norm in quadratic form:

$$\mathbf{B}_{l_2}(\mathcal{L}) = \sum_{(i,j) \in \mathcal{E}} B_{ij}^2(\mathcal{L}_i - \mathcal{L}_j)^2 = \mathcal{L}^T \widetilde{\mathbf{W}} \mathcal{L} \quad (9)$$

where $\widetilde{\mathbf{W}} = \text{diag}(\bar{w}) + \text{diag}(\hat{w}) - 2\mathbf{W}$, $\bar{w}_j = \sum_{j'} w_{jj'}$, $\hat{w}_{j'} = \sum_j w_{jj'}$ and $\mathbf{W} = [w_{jj'}] = [E_{jj'} B_{jj'}]$. The full derivation of the above is included in the supplementary material.

A quadratic optimization problem is NP-hard if the matrix in the quadratic term is non-definite, i.e. the optimization is non-convex. In fact, having even single negative eigenvalue leads to NP-hard problem [20]. Regarding the convexity of the formulation, we have the following proposition.

Proposition 1. *The matrix $\widetilde{\mathbf{W}}$ in Eq.(9) is positive semi-definite.*

The proof is included in the supplementary material. Since $\widetilde{\mathbf{W}}$ is positive semi-definite, the formulation is convex. It is also possible to ensure the matrix to be positive definite

by adding a small positive value to the diagonals. In addition to the well-posedness of this formulation, we show that positive definiteness of the matrix $\widetilde{\mathbf{W}}$ allows the problem to be linearized.

Our linear relaxation is based on the following facts:

$$\mathcal{L}^T \widetilde{\mathbf{W}} \mathcal{L} = \mathcal{L}^T \mathbf{U}^T \mathbf{U} \mathcal{L} = \|\mathbf{U} \mathcal{L}\|_{l_2}^2, \quad (10)$$

where \mathbf{U} is an upper triangular matrix of the same dimension of $\widetilde{\mathbf{W}}$ and $\widetilde{\mathbf{W}} = \mathbf{U}^T \mathbf{U}$ is known as the Cholesky factorization/decomposition. The matrix \mathbf{U} uniquely exists for symmetric positive definite matrix $\widetilde{\mathbf{W}}$.

We observe that the matrix $[\text{diag}(\mathbf{w})\mathbf{D}]$ operating on \mathcal{L} in the l_1 -norm can also be thought of as being factorized from the matrix $\widetilde{\mathbf{W}}$. To see this, we can rewrite Eq. (6) as follows:

$$\begin{aligned} \|\text{diag}(\mathbf{w})\mathbf{D}\mathcal{L}\|_{l_2}^2 &= \mathcal{L}^T [\text{diag}(\mathbf{w})\mathbf{D}]^T [\text{diag}(\mathbf{w})\mathbf{D}] \mathcal{L} \\ &= \mathcal{L}^T \widetilde{\mathbf{W}} \mathcal{L}. \end{aligned} \quad (11)$$

This motivates us to have the following new reformulation of the boundary term \mathbf{B} :

$$\mathbf{B}_{l_1^+}(\mathcal{L}) = \|\mathbf{U} \mathcal{L}\|_{l_1} \quad (12)$$

Here, we call the above norm to be minimized as the l_1^+ -norm boundary term because it is related to the l_2 -norm boundary term in the same fashion of the conventional l_1 -norm boundary term, and it can be more desirable. The above resemblance between the l_1^+ -norm and the l_1 -norm, also motivates us to define our l_1^+ -norm boundary term as the *quasi-total-variation*, since l_1 -norm boundary term is actually the total variation.

A major difference between the conventional l_1 -norm and our l_1^+ -norm boundary terms is that the linear operator \mathbf{U} is more compact than $[\text{diag}(\mathbf{w})\mathbf{D}]$, giving rise to a more compact LP relaxation.

$$\begin{aligned} \min_{\mathcal{L}, \delta^+} B(\mathcal{L}, \delta^+) &= \mathbf{1}^T \delta^+ \\ \text{s. t. :} \quad & -\delta^+ \preceq \mathbf{U} \mathcal{L} \preceq \delta^+ \\ & \forall i, 0 \leq \mathcal{L}_i \leq 1, \delta_i^+ \geq 0, \end{aligned} \quad (13)$$

where δ^+ is an additional vector of auxiliary variables used for the linear relaxation and its dimension is N , as the same as \mathcal{L} . Essentially, Eq.(13) reduces the bound of $\mathbf{U} \mathcal{L}$. The above LP is obtained by applying the equivalence between l_1 -norm minimization and linear programming (Please refer the supplementary materials for full details).

Compared with the conventional LP model in Eq.(3), our model in Eq.(13) has a significantly less number of variables and number of equality or inequality constraints. This is why we call our formulation a compact linear programming relaxation. Specifically, for the image containing N superpixels, there are $N + N \times N$ variables and $N + 2N \times N$ linear constraints for the worst case in the original model [1,16], whereas there are only $2N$ variables and $2N$ linear constraints in our model. The complexity of our model is therefore $O(N^3)$ which is the same as QP according to Eq.(9). The number of variables and constraints does not change when increasing the number of edges in the graph. We will compare the performance of the two formulations experimentally.

4.2 Mathematical relationship between total variation and quasi-total variation

In this subsection, we are particularly interested how tightly the total variation in the form of l_1 -norm can be related to the quasi-total variation in the form of l_1^+ -norm, such that the neat properties of the total variation can be shared by the quasi-total variation.

Let us consider the reduced QR factorization of the rectangular matrix $[\text{diag}(\mathbf{w})\mathbf{D}]$ in the l_1 -norm boundary term, i.e. $[\text{diag}(\mathbf{w})\mathbf{D}] = \mathbf{Q}^{N^2 \times N} \mathbf{R}^{N \times N}$, where \mathbf{Q} is an orthogonal matrix, such that $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}^{N \times N}$, and \mathbf{R} is an upper triangular matrix. The following fact will relate our l_1^+ relaxation to the original l_1 -norm minimization.

Theorem 1. *The upper triangular matrix \mathbf{U} in the l_1^+ -norm minimization model in Eq.(12) is identical to the upper triangular matrix \mathbf{R} in the QR factorization of $[\text{diag}(\mathbf{w})\mathbf{D}]$ in the l_1 -norm minimization model in Eq.(4)*

The proof of this theorem is presented in the supplementary material. The above theorem implies several additional relationships between the l_1 -norm and the l_1^+ -norm.

Corollary 1. *If we replace the l_1 -norm with l_2 -norm, we will have $\|\mathbf{U}\mathcal{L}\|_{l_2} = \|\text{diag}(\mathbf{w})\mathbf{D}\mathcal{L}\|_{l_2}$.*

In other words, \mathbf{U} is an equivalent linear operator of $\text{diag}(\mathbf{w})\mathbf{D}$ for l_2 -norm.

Corollary 2. $\mathbf{U}\mathcal{L} = \mathbf{Q}^T \mathbf{Q} \mathbf{U} \mathcal{L} = \mathbf{Q}^T [\text{diag}(\mathbf{w})\mathbf{D}\mathcal{L}]$.

The above equality implies that the l_1^+ -norm is the l_1 -norm of the linearly transformed weighted gradients, and the transformation matrix is \mathbf{Q} . The weighted variations in \mathcal{L} are projected on the subspace of \mathbf{Q} before calculating the total. Hence, we may also view our l_1^+ -norm as a total subspace-variation. This observation implies that the quasi-total variation minimization may share the discontinuity preservability of the total variation minimization.

Besides, Theorem 1 offers us a stronger relationship between the two formulations in terms of a tight equivalence-of-norm bound.

Theorem 2. *The difference of l_1^+ -norm and l_1 -norm satisfies the following inequalities:*

$$\frac{1}{\|\mathbf{Q}\|_{l_1}} \|\text{diag}(\mathbf{w})\mathbf{D}\mathcal{L}\|_{l_1} \leq \|\mathbf{U}\mathcal{L}\|_{l_1} \leq \|\mathbf{Q}^T\|_{l_1} \|\text{diag}(\mathbf{w})\mathbf{D}\mathcal{L}\|_{l_1} \quad (14)$$

The proof of this theorem is included in the supplementary material.

Comparing this above norm equivalence with Eq.(8), we can conclude that the worst-case difference between the l_1^+ -norm and the l_1 -norm can be much smaller than that between l_2 -norm and l_1 -norm. This is because we are allowed to reduce $\|\mathbf{Q}\|_{l_1}$ and $\|\mathbf{Q}^T\|_{l_1}$ in order to improve the approximation, while the approximation in Eq.(8) is hardy. In addition, according to Theorem 2, we consider our compact LP model in (13) as a tight relaxation of the original l_1 -norm boundary term.

5 Experiments

In the experiment, we will evaluate our model for the boundary term. This evaluation is possible since the seed-initialized interactive object segmentation is generally formulated with only the boundary term, and the seed points (seeds) are incorporated as the hard constraints encoded by the unary potential [2]. We compare our LP for l_1^+ -norm minimization with the original graph cuts (GC) [3], the l_1 -norm minimization via LP [16], and the l_2 -norm minimization by QP [8,24].

5.1 Experimental settings

Data and performance measure We mainly experiment on two segmentation datasets. The first one will be the clownfish segmentation dataset constructed by the authors which contains 62 images of clownfish. The other is the Oxford interactive segmentation benchmark dataset [10]. Ground truth results and user input seeds on objects and backgrounds are provided in both datasets. The clownfish dataset may be simpler than the Oxford dataset because it contains less variations of objects. It can be considered as a controlled dataset, and the Oxford dataset is more like a natural dataset. We shall use the clownfish dataset for proof of concept, and then validate our model under more general situations in the Oxford dataset. The performance of the methods is measured by the overlapping ratio between the labeled region and the ground truth object region:

$$\Gamma = \frac{\text{size}(\text{Result Region} \cap \text{True Region})}{\text{size}(\text{Result Region} \cup \text{True Region})}.$$

To evaluate the performance gain in terms of computation. We perform the conventional LP and our proposed LP on GPU for synthetic data. In this experiment, we randomly generate the model parameters and apply the interior point method to solving the LP. We are unable to experiment on images due to the limit on our hardware.

Implementation issues We adopt superpixelization [15] as a preprocessing to reduce the computational cost. The number of superpixels is around 800 for all test images. We choose the average color of each superpixel to represent the superpixel. We implement all the methods in MATLAB. We used the `linprog` function and `quadprog` function. We use default option settings of the functions. The graph cuts is based on Michael Rubinstein’s implementation¹. There are some parameters in the model for segmentation. We used $c = 0.00001$, $\lambda = 10$ in all the experiments. The threshold value for converting the continuous labels to binary labels is empirically chosen as 0.08. We also experiment on the effect of different threshold values. We perform the experiments on a PC with Intel Core i5-450M (2.4GHz) processor and 4GB memory.

¹ <http://www.mathworks.com/matlabcentral/fileexchange/21310-maxflow>

5.2 Results

The clownfish dataset. We first present and analyze the experimental results for the clownfish dataset. See Fig. 2 for example segmentation results and input seeds (refer supplementary materials for additional results). In addition to the manually drawn background seeds, we include the points at the image border as the background seeds in this experiment. From the results, we can see that the results of the conventional LP is very similar to those by graph cuts as expected. A characteristic of them is that they suffer from the small-cut problem. In contrast, QP may produce larger regions due to the possible diffusion of labels at the boundaries. Thus, the resultant regions can be larger than the desired region. Our method compromises the two types of methods and the overall results may outperform the others, e.g., when LP suffers from small-cut problem and/or QP suffers from large-cut problem. We also present the label map of conventional LP, QP and our LP in Fig. 3. As expected, the solutions of LP are binary without thresholding, and the solutions of QP can be over-smooth. The boundaries in the solutions of our LP are clearer than QP, and the solutions are smoother than LP. Quantitative segmentation results of the clownfish dataset are shown in Fig. 6. The results show that QP slightly outperforms the conventional LP on this dataset, and our method slightly outperforms the others. From Table. 1, we can see that the computational cost of our compact LP model is comparable to QP and requires significantly less computational expenses compared to conventional LP.

The Oxford dataset. After proof of concept by the clown fish dataset, we validate our model on the Oxford dataset. The user input seeds provided in this dataset are generally insufficient for producing a satisfactory segmentation. We adopt the robotuser [10] to simulate the additional user interactions. By increasing the number of interactions, the segmentation results can finally become satisfactory. The maximum number of user interactions is set to 20 in our experiments. See Fig. 4 for example results and supplementary materials for additional results. We can observe that GC and LP performs quite alike, while QP may produce larger regions. In most of the situations our methods produce more accurate segmentation results than QP. We present the solutions of QP and our method before thresholding in Fig. 5. The LP produces binary labels as expected, the QP produces smooth labels near the object boundaries and our method produces piecewise smooth labels with relatively clear discontinuities at the boundaries. The quantitative results are shown as red boxes in Fig. 6. The statistics of the computational costs are shown in Table 1. Very recently, a fast optimization approach has been proposed for solving a similar segmentation model [27]. However, the computational cost of their approach for 760 superpixels is 2 times of our cost on a PC better than ours.

To quantitatively reveal the effect of the discontinuity preservability of our method, we further consider the robustness of the segmentation to threshold values. We hypothesize that the continuous labels with clear discontinuities at the boundaries will be robust to different threshold values. Therefore, we generate a vector of 100 threshold values equally spaced in $[0, 1]$ for the evaluation. We apply all these threshold values to the continuous labels of QP and our method. Surprisingly, we observe that our method overwhelmingly outperforms the QP for almost all the threshold values in the sense of

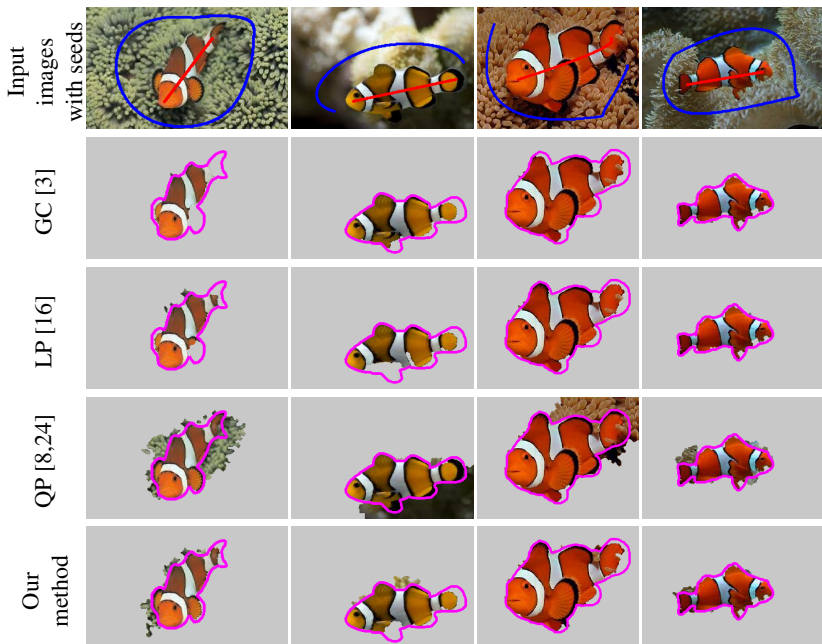


Fig. 2: Example results of the seed-initialized interactive segmentation on clownfish dataset. The results are shown as extracted image regions against the ground truth shape contours in purple.

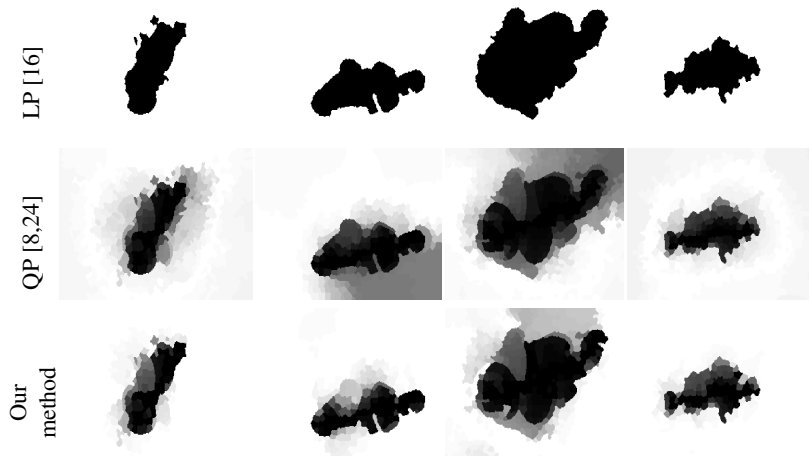


Fig. 3: Example labels of segmentation results in Fig. 2.

average overlapping ratio. See Fig. 7 for the plots of mean performance with standard deviation.



Fig. 4: Comparison of segmentation performance on Oxford dataset. The top row show the input images overlaid with input seeds. The bottom rows show extracted image regions against the ground truth shape contours in purple.

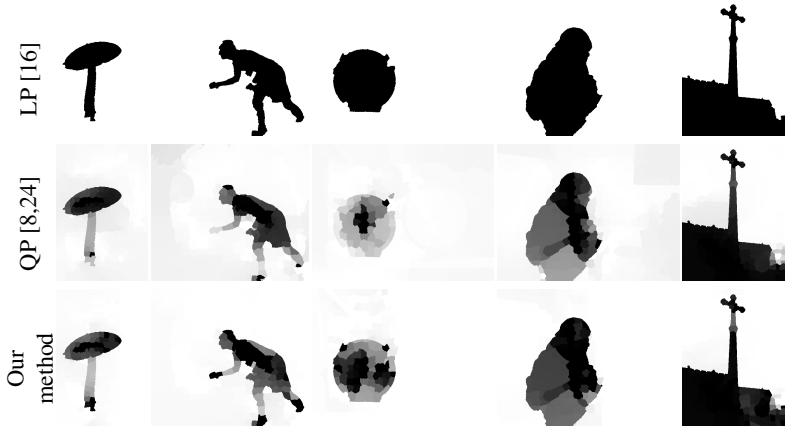


Fig. 5: Continuous labels before thresholding from LP, QP and our method on example inputs in Fig. 4

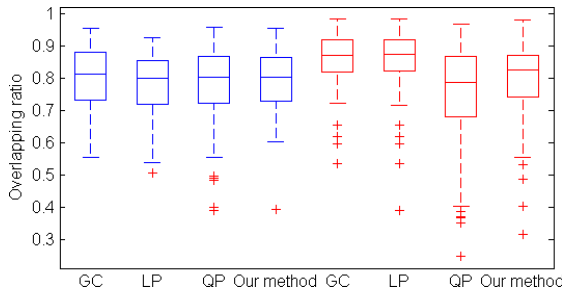


Fig. 6: Quantitative results of the experiments. Blue boxes are the results for clownfish dataset and the red boxes are the results for Oxford dataset.

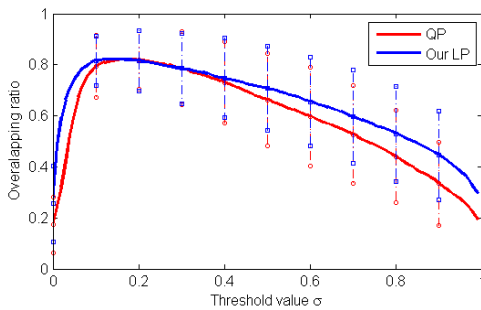


Fig. 7: Comparison of QP and our LP with different threshold values on the entire Oxford dataset.

Table 1: Comparison of computational costs.

	LP [16]	QP [8,24]	Our method
Worst-case complexity	$O(N^6)$	$O(N^3)$	$O(N^3)$
Time (s)	72.35 ± 9.33	1.13 ± 0.48	12.9 ± 2.61

6 Conclusion

In this paper, we proposed a novel LP relaxation for the binary sub-modular MRF model. Our LP relaxation is based on a novel l_1^+ -norm minimization in this paper, and it contains significantly fewer variables and constraints compared to the conventional LP. We also show that our l_1^+ -norm minimization is tightly related to the total variation minimization, according to which we argue that the discontinuities in the solution at the object boundaries can be well preserved. Experimental results show that our method is significantly faster than the conventional LP. Besides, given the same order of computational complexity, our method uniformly outperforms the QP when converting the continuous labels to binary labels using various threshold values on the entire Oxford dataset. Our model may be of use to many other problems modeled by MRF.

References

1. Arvind Bhusnurm and Camillo J Taylor. Graph cuts via l_1 norm minimization. *TPAMI*, 30(10):1866–1871, 2008.
2. Yuri Boykov and Marie-Pierre Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In *ICCV*, 2001.
3. Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *TPAMI*, 23(11):1222–1239, November 2001.
4. T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *ECCV*, volume 3024, pages 25–36, 2004.
5. Antonin Chambolle. Total variation minimization and a class of binary mrf models. In *Energy minimization methods in computer vision and pattern recognition*, pages 136–152. Springer, 2005.
6. T.F. Chan and L.A. Vese. Active contours without edges. *TIP*, 10(2):266–277, 2001.
7. Shuai Che, Michael Boyer, Jiayuan Meng, David Tarjan, Jeremy W Sheaffer, and Kevin Skadron. A performance study of general-purpose applications on graphics processors using cuda. *Journal of parallel and distributed computing*, 68(10):1370–1380, 2008.
8. Leo Grady. Random walks for image segmentation. *TPAMI*, 28(11):1768–1783, 2006.
9. D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)*, 51(2):271–279, 1989.
10. V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman. Geodesic star convexity for interactive image segmentation. In *CVPR*, 2010.
11. Jorg H Kappes, Bjoern Andres, Fred A Hamprecht, Christoph Schnorr, Sebastian Nowozin, Dhruv Batra, Sungwoong Kim, Bernhard X Kausler, Jan Lellmann, Nikos Komodakis, et al.

- A comparative study of modern inference techniques for discrete energy minimization problems. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1328–1335. IEEE, 2013.
12. M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321–331, 1988.
 13. Vladimir Kolmogorov and Carsten Rother. Minimizing nonsubmodular functions with graph cuts—a review. *TPAMI*, 29(7):1274–1279, 2007.
 14. Nikos Komodakis, Nikos Paragios, and Georgios Tziritas. Mrf energy minimization and beyond via dual decomposition. *TPAMI*, 33(3):531–552, 2011.
 15. Alex Levinstein, Adrian Stere, Kiriakos N. Kutulakos, David J. Fleet, Sven J. Dickinson, and Kaleem Siddiqi. Turbopixels: Fast superpixels using geometric flows. *TPAMI*, 31:2290–2297, December 2009.
 16. Hongdong Li and Chunhua Shen. Interactive color image segmentation with linear programming. *Machine Vision and Applications*, 21(4):403–412, 2010.
 17. Stan Z. Li. *Markov random field modeling in image analysis*. Springer-Verlag New York, Inc., 3rd edition, 2009.
 18. Jiangyu Liu and Jian Sun. Parallel graph-cuts by adaptive bottom-up merging. In *CVPR*, pages 2181–2188. IEEE, 2010.
 19. Nimrod Megiddo. Linear programming in linear time when the dimension is fixed. *Journal of the ACM (JACM)*, 31(1):114–127, 1984.
 20. Panos M. Pardalos and Stephen A. Vavasis. Quadratic programming with one negative eigenvalue is NP-hard. *Journal of Global Optimization*, 1:15–22, 1991.
 21. Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259 – 268, 1992.
 22. Thomas Schoenemann, Fredrik Kahl, Simon Masnou, and Daniel Cremers. A linear framework for region-based image segmentation and inpainting involving curvature penalization. *IJCV*, 99(1):53–68, 2012.
 23. Alexander Shekhovtsov and Václav Hlaváč. A distributed mincut/maxflow algorithm combining path augmentation and push-relabel. *IJCV*, 104(3):315–342, 2013.
 24. Ali Kemal Sinop and Leo Grady. A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm. In *CVPR*. IEEE, 2007.
 25. Jian Sun, Nan-Ning Zheng, and Heung-Yeung Shum. Stereo matching using belief propagation. *TPAMI*, 25(7):787–800, July 2003.
 26. Lloyd N Trefethen and David Bau III. *Numerical linear algebra*. Number 50. Siam, 1997.
 27. Peng Wang, Chunhua Shen, and Anton van den Hengel. A fast semidefinite approach to solving binary quadratic problems. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR’13)*, Oregon, USA, 2013.
 28. Yinyu Ye. An $o(n^3l)$ potential reduction algorithm for linear programming. *Mathematical programming*, 50(1-3):239–258, 1991.

A Appendix

In this supplementary material, we include the lengthy proofs, derivations and additional experimental results that we omitted in the paper due to the page limit.

To evaluate the performance gain in terms of computation, we perform the conventional LP and our proposed LP on matlab GPU for synthetic data. In this experiment, we randomly generate the model parameters for the Eq. (3), we factorize the boundary term to arrive at Eq. (13) and we apply the interior point method to solving both the LP problems. We are unable to experiment on images due to the limit on our hardware. Our graphics card is the mobile NVIDIA Geforce 780M with 3GB graphics memory on our laptop. It thus does not allow the GPU implementation for the conventional LP on images containing hundreds of superpixels. Let's consider 200 superpixels in an image. We will have a constraint matrix containing $(200 + 200 \times 200) \times (200 + 2 \times 200 \times 200) = 3224040000$ elements. Since the matlab GPU only supports full matrix, we will need 24 GB graphics memory for storing this full matrix, which is frankly impossible.

The average computational costs on GPU are shown in Fig. 8. We also plot the baseline computational cost of LP on CPU. It can be seen that the computational cost of LP on GPU tends to be lower than that of the LP on CPU when the number of variables increase. It is interesting to see that the computational cost of our LP on GPU remains steady w.r.t. the number of variables.

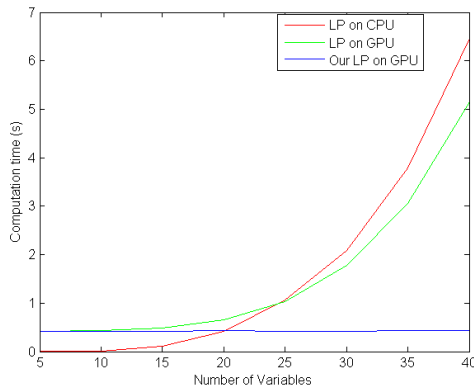


Fig. 8: Comparison of computational times on GPU.

A.1 Proofs and derivations

Derivation of Eq. (9)

$$\begin{aligned}
\mathbf{B}(\mathcal{L}) &= \sum_{j,j'} E_{jj'} B_{jj'} |\mathcal{L}_j - \mathcal{L}_{j'}|^2 \\
&= \sum_{j,j'} w_{jj'} (\mathcal{L}_j^2 + \mathcal{L}_{j'}^2 - 2\mathcal{L}_j \mathcal{L}_{j'}) \\
&= \sum_j \mathcal{L}_j^2 \sum_{j'} w_{jj'} + \sum_{j'} \mathcal{L}_{j'}^2 \sum_i w_{jj'} \\
&\quad - 2 \sum_{j,j'} w_{jj'} \mathcal{L}_j \mathcal{L}_{j'} \\
&= \sum_j \mathcal{L}_j^2 \bar{w}_j + \sum_{j'} \mathcal{L}_{j'}^2 \hat{w}_{j'} - 2 \sum_{j,j'} w_{jj'} \mathcal{L}_j \mathcal{L}_{j'} \\
&= \mathcal{L}^T \widetilde{\mathbf{W}} \mathcal{L}
\end{aligned} \tag{A-1}$$

where $\widetilde{\mathbf{W}} = \text{diag}(\bar{w}) + \text{diag}(\hat{w}) - 2\mathbf{W}$, $\mathbf{W} = [w_{jj'}] = [E_{jj'} B_{jj'}]$.

Proof of Proposition 4.1

Proof. The definition of $\widetilde{\mathbf{W}}$ is as follows.

$$\widetilde{\mathbf{W}} = \text{diag}(\bar{w}) + \text{diag}(\hat{w}) - 2\mathbf{W} \tag{A-2}$$

where

$$\bar{w}_j = \sum_k w_{jl} = \sum_k w_{lj'} = \hat{w}_{j'}, \text{ if } j = j'. \tag{A-3}$$

In short $\text{diag}(\bar{w}) = \text{diag}(\hat{w})$. Note that $w_{jj'} = 0$ for $j = j'$. Hence, we have the following.

$$\widetilde{\mathbf{W}}_{jj'} = \begin{cases} 2\bar{w}_j, & \text{for } j = j' \\ -2w_{jj'}, & \text{otherwise} \end{cases} \tag{A-4}$$

Therefore, matrix $\widetilde{\mathbf{W}}$ is a symmetric diagonal dominant matrix, and the diagonal elements are nonnegative. Such matrix is a positive semi-definite matrix. \square

Equivalence between L_1 norm minimization and linear programming

$$\begin{aligned}
\sum_i |y_i| &= \min_{y^+} \sum_i y_i^+ \\
\text{s.t. } &\forall i, -y_i^+ \leq y_i \leq y_i^+ \\
&\forall i, y_i^+ > 0.
\end{aligned} \tag{A-5}$$

Proof of Theorem 4.2

Proof. Substituting $[\text{diag}(\mathbf{w})\mathbf{D}] = \mathbf{Q}^{N^2 \times N} \mathbf{R}^{N \times N}$ into Eq.(4), we obtain the following form of the boundary term.

$$\mathbf{B}_{l_1}(\mathcal{L}) = \|\mathbf{Q}\mathbf{R}\mathcal{L}\|_{l_1} \quad (\text{A-6})$$

where we applied the QR factorization. The l_2 relaxation of this form will lead to

$$\begin{aligned} \mathbf{B}_{l_2}(\mathcal{L}) &= (\mathcal{L}^T \mathbf{R}^T \mathbf{Q}^T \mathbf{Q} \mathbf{R} \mathcal{L})^{1/2} \\ &= (\mathcal{L}^T \mathbf{R}^T \mathbf{R} \mathcal{L})^{1/2} \\ &= \|\mathbf{R}\mathcal{L}\|_{l_2} \end{aligned} \quad (\text{A-7})$$

The corresponding l_1^+ -norm minimization is therefore the following

$$\mathbf{B}_{l_1^+}(\mathcal{L}) = \|\mathbf{R}\mathcal{L}\|_{l_1} \quad (\text{A-8})$$

Note that the Cholesky decomposition is unique and \mathbf{R} is upper-triangular. We can conclude that $\mathbf{U} = \mathbf{R}$. \square

Proof of Theorem 4.5

Proof. We prove the left hand side first.

$$\begin{aligned} \|\text{diag}(\mathbf{w})\mathbf{D}\mathcal{L}\|_{l_1} &= \|\mathbf{Q}\mathbf{U}\mathcal{L}\|_{l_1} \\ &\leq \|\mathbf{Q}\|_{l_1} \|\mathbf{U}\mathcal{L}\|_{l_1} \\ &\Leftrightarrow \\ \frac{1}{\|\mathbf{Q}\|_{l_1}} \|\text{diag}(\mathbf{w})\mathbf{D}\mathcal{L}\|_{l_1} &\leq \|\mathbf{U}\mathcal{L}\|_{l_1} \end{aligned} \quad (\text{A-9})$$

where we have replaced \mathbf{R} with \mathbf{U} . The right hand side is likewise.

$$\begin{aligned} \|\mathbf{U}\mathcal{L}\|_{l_1} &= \|\mathbf{Q}^T \mathbf{Q} \mathbf{U} \mathcal{L}\|_{l_1} \\ &\leq \|\mathbf{Q}^T\|_{l_1} \|\mathbf{Q} \mathbf{U} \mathcal{L}\|_{l_1} \\ &= \|\mathbf{Q}^T\|_{l_1} \|\text{diag}(\mathbf{w})\mathbf{D}\mathcal{L}\|_{l_1}, \end{aligned} \quad (\text{A-10})$$

which completes the proof. \square

A.2 Extra experimental results

We present some additional experimental results in Figs. 9-11. The corresponding continuous solutions before thresholding are presented in Figs. 12-14.

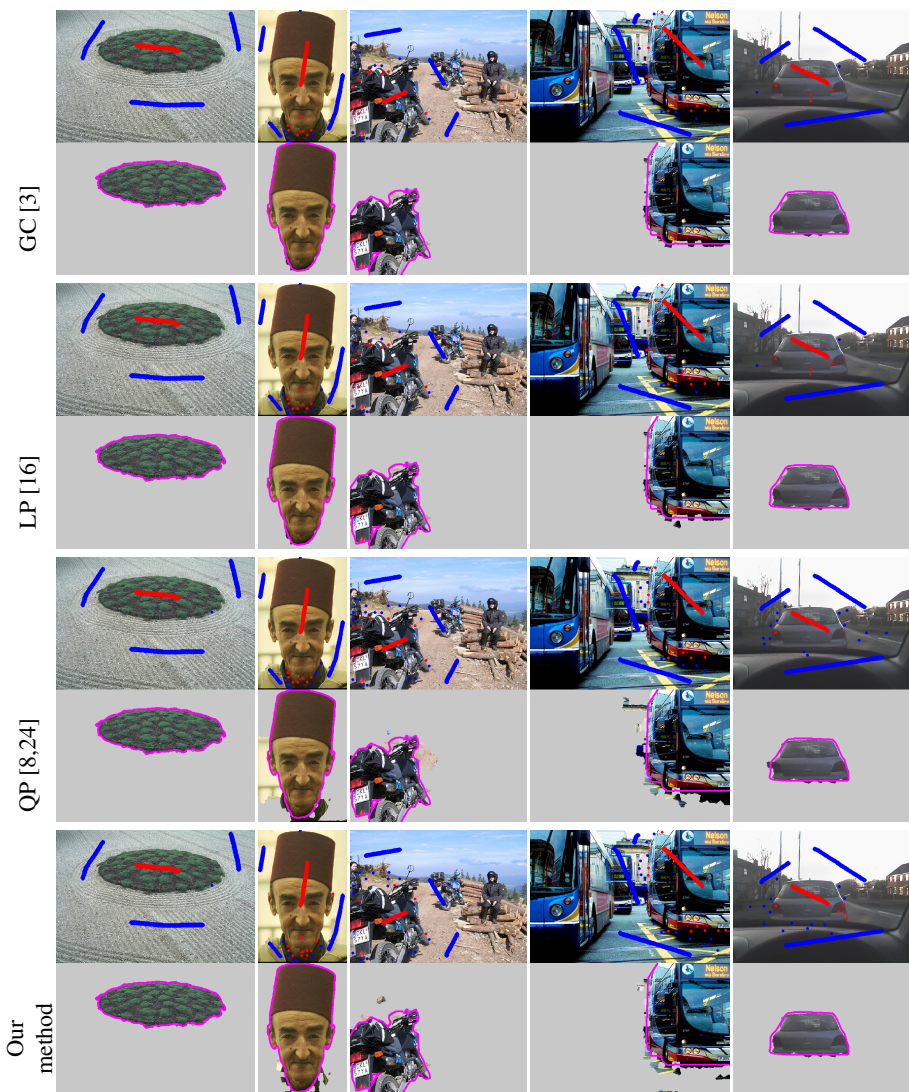


Fig. 9: Additional experimental results of segmentation on Oxford dataset-I. The input and result images are shown as top-bottom pairs. The top rows show the input images overlaid with input seeds. The bottom rows show extracted image regions against the ground truth shape contours in purple.



Fig. 10: Additional experimental results of segmentation on Oxford dataset-II

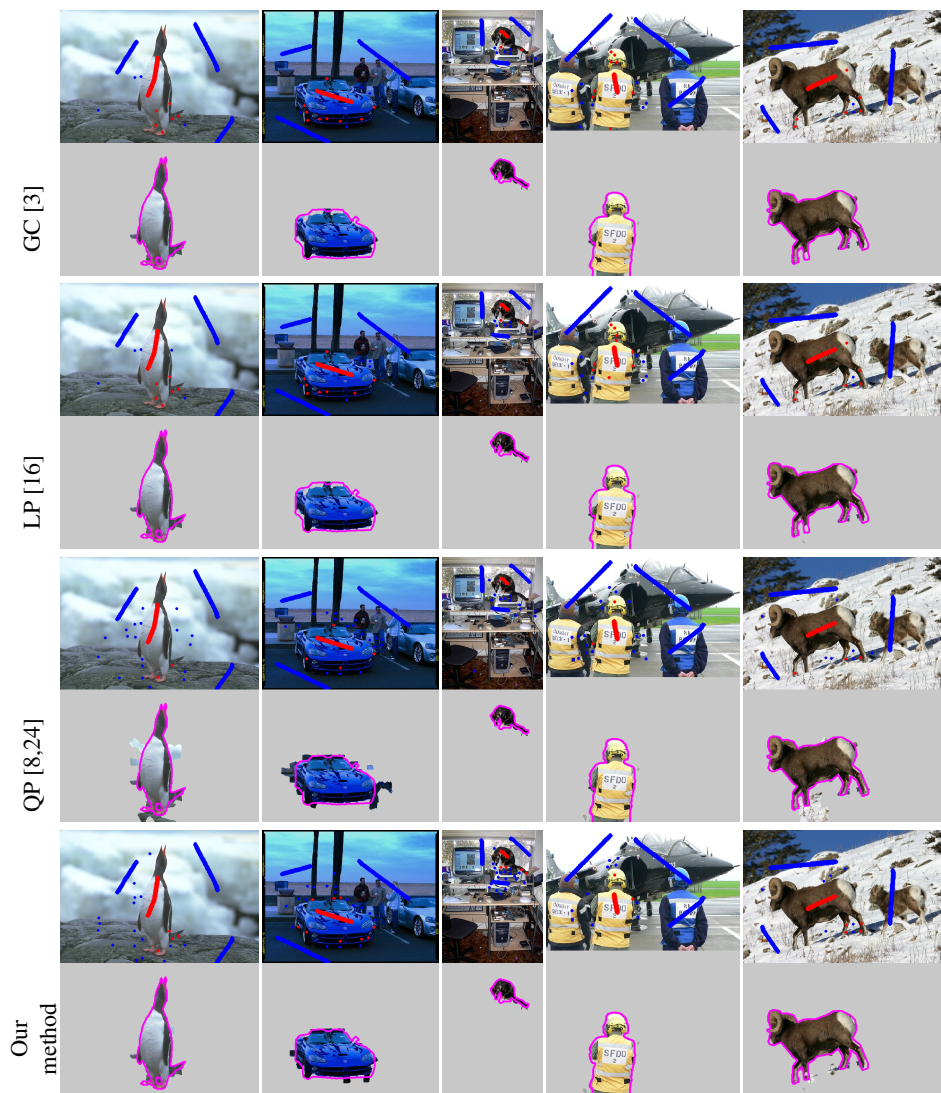


Fig. 11: Additional experimental results of segmentation on Oxford dataset-III

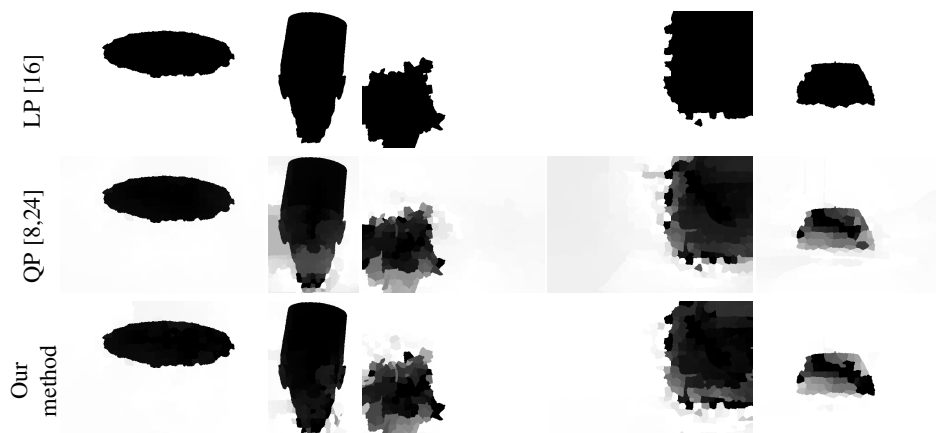


Fig. 12: Continuous labels before thresholding from LP, QP and our method on example inputs in Fig. 9

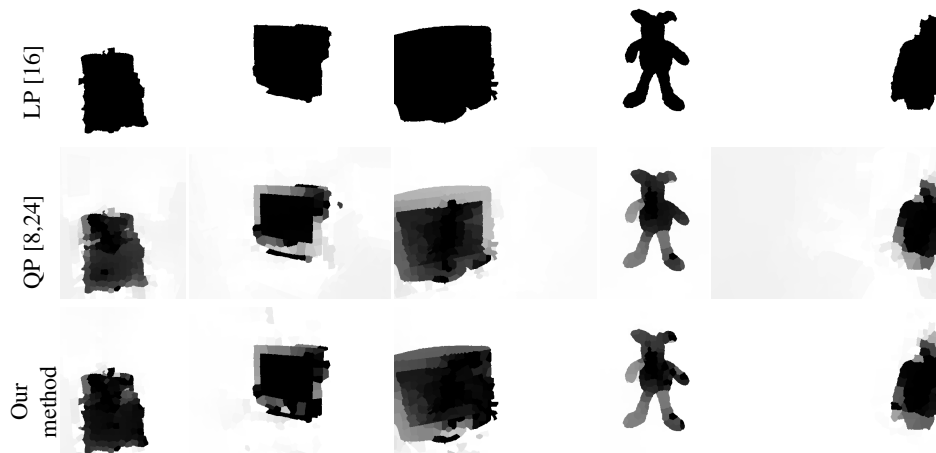


Fig. 13: Continuous labels before thresholding from LP, QP and our method on example inputs in Fig. 10

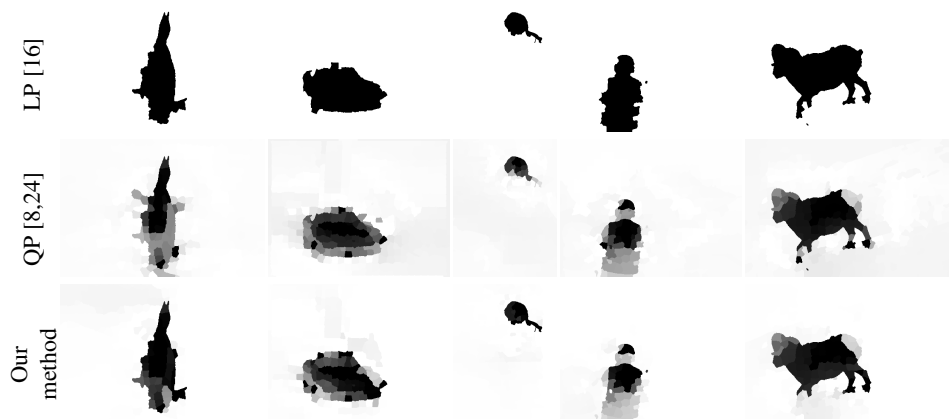


Fig. 14: Continuous labels before thresholding from LP, QP and our method on example inputs in Fig. 11